

Data Lakehouse

State of Arizona – ADHS

Project Investment Justification (PIJ)

January 18, 2023



ARIZONA DEPARTMENT
OF HEALTH SERVICES

Agency Vision

Health and Wellness for all Arizonans

Agency Mission

*To promote, protect, and improve the health and wellness of individuals
and communities in Arizona.*



Team Introduction

Roles Present at ITAC

- Susan Robinson - Chief Business Intelligence Officer
- Paula Mattingly - Chief Information Officer
- Luke Evans - IT Project Management Office Manager

Project Introduction

Stated Operational/Business Issue

- Lack of centrally governed data analytics and dataset repository
- Slows down information sharing with partners
- Increased probability of data errors
- No single version of the truth when determining or reporting on public health
 - Anomalies
 - Epidemics
 - Pandemics

Benefit to the State Agency and Constituents

- Increased agility to respond to health emergencies
- Increased confidence in fact-based data analytics
- Enhances data management capabilities
- Streamlines access and enhances data quality and systems
- Aligns with and supports the Pandemic Recovery and Resiliency Plan

Note - The scope of this project request is to stand up the new Data Lakehouse architecture and ingest mortality data. Subsequent efforts will ingest other data sets.

Proposed Solution

Overview of Proposed Solution

1. **Procure two best-of-breed data lakehouse tools:**
 - a. **Snowflake** - Cloud-based data lakehouse repository
 - b. **Matillion** - Cloud-based data ingestion
2. **Select a Data Lakehouse professional services firm**
3. **Analyze, create, and practice Data Lakehouse governance best practices**
4. **Build out the architecture**
 - a. **Snowflake** will reside on the Snowflake AWS cloud
 - b. **Matillion** will reside on the ADHS AWS cloud
5. **Create Matillion routines to ingest ADHS mortality data into Snowflake and transform data into views required by users.**
6. **Securely allow internal users group to ingest mortality data. Monitor for access anomalies.**

Project Responsibilities

Identify Proposed Solutions Responsibilities

Agency

1. Project & vendor management
2. Cloud infrastructure support
3. Security assessments
4. Tool/data access management
5. User acceptance

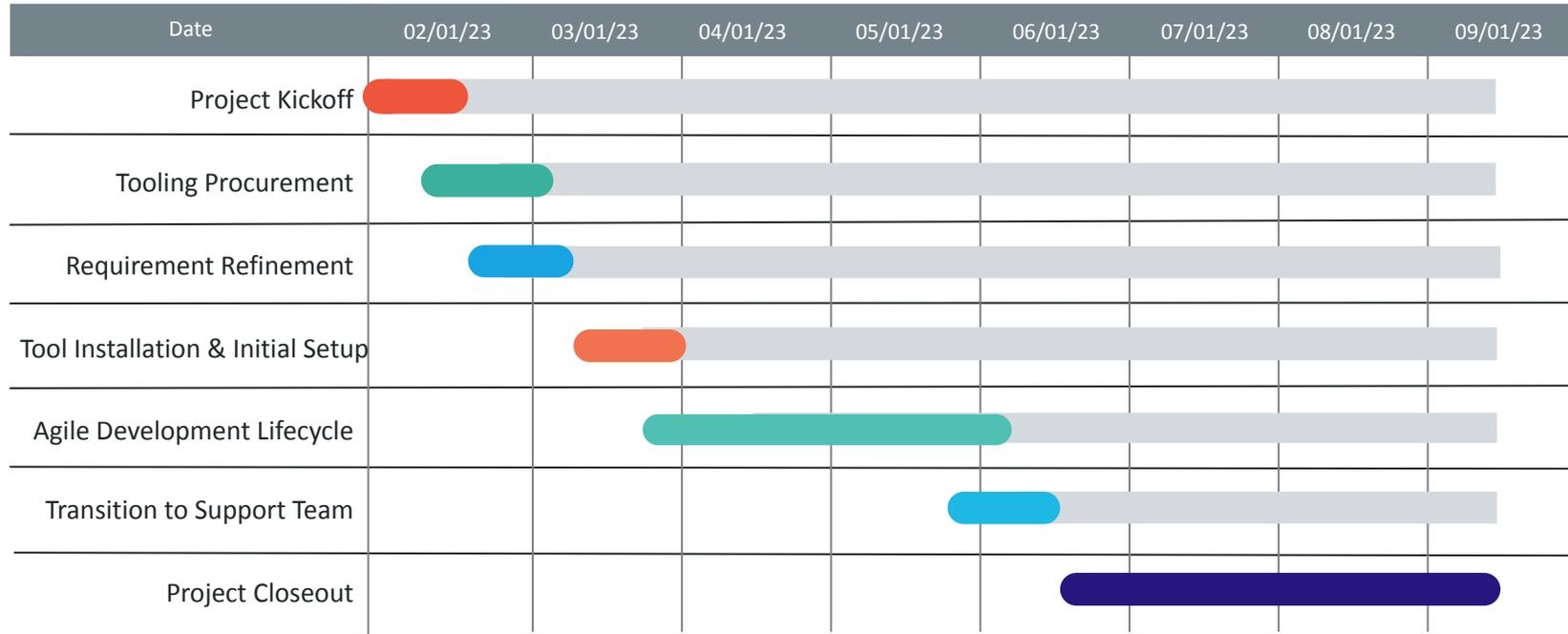
Shared

1. Governance creation for
 - a. Data retention
 - b. Data access security
 - c. Code management
 - d. Change management
2. Architecture design
3. Backlog grooming
4. Technical tool training and knowledge transfer
5. End User training

Vendor/Contractor

1. Development Lifecycle Management (Agile)
2. Tooling installation
3. User requirements facilitation
4. Quality assurance testing
5. User acceptance facilitation

Project Timeline



Project Costs

Project Costs by Category	FY23	FY24	FY25	FY26	FY27	Total
Professional & Outside Services (Contractors)	\$841,976	\$0	\$0	\$0	\$0	\$841,976
Hardware	\$0	\$0	\$0	\$0	\$0	\$0
Software	\$79,624	\$78,000	\$78,000	\$78,000	\$78,000	\$391,624
Communications	\$0	\$0	\$0	\$0	\$0	\$0
Facilities	\$0	\$0	\$0	\$0	\$0	\$0
License & Maintenance Fees	\$28,986	\$21,511	\$21,511	\$21,511	\$21,511	\$115,030
Other Operational Expenditures	\$0	\$0	\$0	\$0	\$0	\$0
Total Development	\$950,586	\$0	\$0	\$0	\$0	\$950,586
Total Operational	\$0	\$99,511	\$99,511	\$99,511	\$99,511	\$398,044

What Success Looks Like

Change Management

- a. Project Milestones
 - a. Communications
 - b. Training
 - c. Other CM Actions

Measures of Success

- a. 100% of current analytics and reports based on the ADHS mortality data have been successfully migrated to use Lakehouse data within 90 days of the final release
- b. 90% of current ADHS end users of the data are able to access the data within the Data Lakehouse according to their specified access permissions, within 30 days of the final release.

Proposed Solution

Due Diligence and Method of Procurement

1. *11 State-authorized professional services vendors were requested to draft proposals*
2. *4 vendors provided responses*
3. *Vendors were rated on their experience and ability to execute per the project and Agency parameters*
4. *The selected vendor is on State contract*

Technology

1. *A thorough Discovery phase was executed to review Data Lakehouse products that would best meet immediate and future needs.*
2. *Major criteria categories for selecting the vendors was*
 - a. *Performance, scalability, and reliability*
 - b. *Data governance and security*
 - c. *Data sharing*
 - d. *Time to value*
 - e. *Ease of use*

Q & A Session

Appendix

Program Structure



Financial Impact

Overview of Current Budget Position

State Funding			
APF			
	Appropriated	Favorably Reviewed	ADOA Transferred
Total FY22 Appropriated	\$	\$	\$
Total FY23 Appropriated	\$	\$	\$
Total FY24 Appropriated	\$	\$	\$
Total FY25 Appropriated	\$	\$	\$
Total APF	\$	\$	\$
General or Non-APF State Fund			
Fiscal Year	Planned		
FY22	\$		
FY23	\$		
Total General Fund	\$		
Total of All State Funding	\$		
Federal Funding			
	Appropriated	Approved to Spend	Available to Spend
Name/Number of Fund	\$1,301,010	\$1,301,010	\$1,301,010
Actuals			
APF/GF Spent		Federal Spent	
Name/Number of Fund	\$	Name/Number of Fund	\$
Name/Number of Fund	\$	Name/Number of Fund	\$
Total APF/GF Spent	\$	Total Federal Funds Spent	\$
Total Fund Remaining			
Total Funds Available to Spend	\$1,301,010		
Total Funds Spent	\$0		
Total Funds Remaining	\$1,301,010		

- *This project is 100% federally funded.*
- *Funding for this project expires July 1, 2024*